

# Fast and Robust Bootstrap

Matías Salibián-Barrera<sup>1</sup>, Stefan Van Aelst<sup>2\*</sup>, Gert Willems<sup>2</sup>

<sup>1</sup> University of British Columbia, Dept. of Statistics, Vancouver, Canada.

<sup>2</sup> Ghent University - UGent, Dept. of Applied Mathematics and Computer Science, Gent, Belgium.

Received: date

**Abstract** In this paper we review recent developments on a bootstrap method for robust estimators which is computationally faster and more resistant to outliers than the classical bootstrap. This fast and robust bootstrap method is, under reasonable regularity conditions, asymptotically consistent. We describe the method in general and then consider its application to perform inference based on robust estimators for the linear regression and multivariate location-scatter models. In particular, we study confidence and prediction intervals and tests of hypotheses for linear regression models, inference for location-scatter parameters and principal components, and classification error estimation for discriminant analysis.

**Key words** Robust inference, bootstrap, multivariate models

---

\* The research of Stefan Van Aelst was supported by a grant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen).

## 1 Introduction

Inference based on robust estimators for linear regression or multivariate location-scatter models is generally based on the asymptotic distribution of these estimators. These asymptotic distributions have, until recently, been studied mainly under the assumption that the distribution of the data is symmetric or elliptical. For inference purposes one can, for example, use the normal approximation and an estimate of the appropriate asymptotic variance-covariance matrix. Unfortunately, these assumptions (symmetric errors for the linear regression model, or elliptical distributions for multivariate observations) do not hold in most practical situations when robust methods would be recommended, i.e. when outliers may be present in the data.

If one relaxes the assumption on symmetric or elliptical distributions, the calculation of the asymptotic distribution of robust estimators becomes involved (for example, for the linear regression case see Carroll, 1978, 1979; Huber, 1981; Rocke and Downs, 1981; Carroll and Welsh, 1988; Salibián-Barrera, 2000; and Croux *et al.*, 2003).

An alternative approach to perform inference based on robust estimators is given by the bootstrap (Efron 1979) which does not generally require stringent distributional assumptions. In most practical cases the bootstrap is based on simulating the distribution of the estimator of interest by generating a large number of new samples randomly drawn from the original data. The estimator is recalculated in each of these “bootstrap samples”, and the empirical distribution of these “bootstrapped estimates” yields an approximation to the estimator’s true sampling distribution. This method has been extensively studied for diverse models.

However, two main problems arise when we want to use the bootstrap to estimate the sampling distribution of robust estimators:

- *numerical instability*: the simulated bootstrap distribution may be severely affected by bootstrap samples with a higher proportion of outliers than in the original data set; and
- *computational cost*: due to the non-convex optimization problems that generally have to be solved in order to calculate robust estimates, it may not be feasible to re-compute these estimators a few thousand times.

Intuitively, the reason for the *numerical instability* mentioned above is as follows. Since outlying and non-outlying observations have the same chance of belonging to any bootstrap sample, with a certain positive probability, the proportion of outliers in a bootstrap sample may be larger than the fraction of contamination that the robust estimator can tolerate. In other words, a certain proportion of the re-calculated values of the robust estimate may be heavily influenced by the outliers in the data. Thus, the tails of the bootstrap distribution can be heavily influenced by the outliers, regardless of the robustness of the statistic being bootstrapped. Stromberg (1997) and Singh (1998) quantified this problem in the context of robust location estimation. The former pointed out that the usual bootstrap estimates for standard errors or covariances have a breakdown point converging to  $1/n$  (with  $n$  the sample size) when the number of bootstrap samples increases. This holds independently of the breakdown point of the bootstrapped estimate. Singh (1998) investigated the estimates of the quantiles of the asymptotic distribution of univariate location estimates. He defined the breakdown point for bootstrap quantiles and showed that it is disappointingly low, although not as low as that of the bootstrap estimator of standard errors. He proposed to draw the bootstrap samples from Winsorized observations and showed that the quantile estimates obtained with this method have the highest attainable breakdown point and that they converge to the quantiles of the asymptotic distribution of the estimate. However, for higher dimensional models (as linear regression), this approach may still be computationally too costly, as it needs to fully re-compute the robust estimators on the bootstrap samples.

The computation of robust estimates is generally not trivial since it typically requires solving a non-convex minimization problem in multiple dimensions. Moreover, the objective function of this optimization problem is often defined implicitly. The number of bootstrap samples needed to obtain reliable distribution estimates naturally grows with the dimension of the statistic, which makes the problem computationally even more expensive to solve. This large number of non-linear optimization problems may render the method unfeasible for high dimensional problems. As an illustration of the computational cost that can be expected, the evaluation of 5000 bootstrap recalculations of a 50% breakdown point and 95% efficient MM-regression estimator on a simulated data set with 2000 observations and 10 explanatory variables took about 2 CPU hours on an IBM PC 3.1GHz with 1GB of RAM. The computations were carried out in R using the `robustbase` library (available at the CRAN website <http://cran.r-project.org>). Note that this implementation of MM-regression estimators includes the fast-S algorithm (Salibián-Barrera and Yohai 2006) for the initial S-estimator.

Recently, Salibián-Barrera and Zamar (2002) introduced a new bootstrap method for robust estimators which has the advantage of being at the same time easy to compute and resistant to outliers in the sample (in other words, it can overcome the two problems mentioned above). The method can in principle be used for any estimator that can be written as the solution of a system of smooth fixed-point equations. Examples of well-known robust estimators that fulfill this condition are S-estimators (Rousseeuw and Yohai 1984; Davies 1987), MM-estimators (Yohai 1987; Tatsuoka and Tyler 2000) and  $\tau$ -estimators (Yohai and Zamar 1988; Lopuhaä 1992).

In what follows we describe the method in general. Its application to linear regression and multivariate models are discussed in detail in Sections 2 and 3, respectively. Consider a sample  $\mathcal{Z}_n$  and let  $\hat{\boldsymbol{\theta}}_n \in \mathbb{R}^m$  be the robust parameter estimates of interest ( $\hat{\boldsymbol{\theta}}_n$  may include estimates for the regression coefficients and scale in the linear model, or for the location and scatter matrix in the multivariate model). Suppose

now that  $\hat{\boldsymbol{\theta}}_n$  can be represented as a solution of fixed-point equations:

$$\hat{\boldsymbol{\theta}}_n = \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n), \quad (1)$$

where the function  $\mathbf{g}_n : \mathbb{R}^m \rightarrow \mathbb{R}^m$  generally depends on the sample  $\mathcal{Z}_n$ . Given a bootstrap sample  $\mathcal{Z}_n^*$ , the recalculated estimate  $\hat{\boldsymbol{\theta}}_n^*$  then solves

$$\hat{\boldsymbol{\theta}}_n^* = \mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n^*), \quad (2)$$

where the function  $\mathbf{g}_n^*$  now depends on  $\mathcal{Z}_n^*$ . As explained above, computing  $\hat{\boldsymbol{\theta}}_n^*$  for every bootstrap sample  $\mathcal{Z}_n^*$  can be a computationally expensive task for robust estimates such as S- or MM-estimates. Moreover, even though we may assume that the solution to (1) was resistant to outliers in  $\mathcal{Z}_n$ , this does not guarantee that we will obtain an equally resistant solution to (2) as  $\mathbf{g}_n^*$  is potentially more severely affected by outliers than  $\mathbf{g}_n$  is.

Instead of computing  $\hat{\boldsymbol{\theta}}_n^*$ , however, we can compute the approximation

$$\hat{\boldsymbol{\theta}}_n^{1*} := \mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n) \quad (3)$$

which can be viewed as one-step estimation of  $\hat{\boldsymbol{\theta}}_n^*$  with initial value  $\hat{\boldsymbol{\theta}}_n$ . It can be shown that, under certain conditions, the distribution of  $\hat{\boldsymbol{\theta}}_n^*$  consistently estimates the sampling distribution of  $\hat{\boldsymbol{\theta}}_n$ . It is intuitively clear, however, that the distribution of  $\hat{\boldsymbol{\theta}}_n^{1*}$  may not have this property in general. Indeed, the recalculated  $\hat{\boldsymbol{\theta}}_n^{1*}$  typically underestimate the actual variability of  $\hat{\boldsymbol{\theta}}_n$ , mainly because every bootstrap sample uses the same initial value in the one-step approximation. To remedy this, a linear correction can be applied as follows. Using the smoothness of  $\mathbf{g}_n$ , we can calculate a Taylor expansion about  $\hat{\boldsymbol{\theta}}_n$ 's limiting value  $\boldsymbol{\theta}$ ,

$$\hat{\boldsymbol{\theta}}_n = \mathbf{g}_n(\boldsymbol{\theta}) + \nabla \mathbf{g}_n(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + R_n, \quad (4)$$

where  $R_n$  is the remainder term and  $\nabla \mathbf{g}_n(\cdot) \in \mathbb{R}^{m \times m}$  is the matrix of partial derivatives. When the remainder term is negligible ( $R_n = o_p(1)$ ), equation (4) can be rewritten as

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \sim [\mathbf{I} - \nabla \mathbf{g}_n(\boldsymbol{\theta})]^{-1} \sqrt{n}(\mathbf{g}_n(\boldsymbol{\theta}) - \boldsymbol{\theta}),$$

where  $\sim$  denotes that both sides have the same limiting distribution. Under certain conditions we will have that  $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \sim \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$  and  $\sqrt{n}(\mathbf{g}_n^*(\boldsymbol{\theta}) - \boldsymbol{\theta}) \sim \sqrt{n}(\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n) - \hat{\boldsymbol{\theta}}_n)$ . If we furthermore approximate  $[\mathbf{I} - \nabla \mathbf{g}_n(\boldsymbol{\theta})]^{-1}$  by  $[\mathbf{I} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}$  we obtain

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \sim [\mathbf{I} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)]^{-1} \sqrt{n}(\mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n) - \hat{\boldsymbol{\theta}}_n). \quad (5)$$

Define then the linearly corrected version of the one-step approximation (3) as

$$\hat{\boldsymbol{\theta}}_n^{R*} := \hat{\boldsymbol{\theta}}_n + [\mathbf{I} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}(\hat{\boldsymbol{\theta}}_n^{1*} - \hat{\boldsymbol{\theta}}_n). \quad (6)$$

When (5) indeed holds,  $\hat{\boldsymbol{\theta}}_n^{R*}$  will be estimating the same limiting distribution as the actual bootstrap calculations  $\hat{\boldsymbol{\theta}}_n^*$  do. For the case of S- and MM-estimates for regression and multivariate location and scatter, formal consistency proofs under relatively mild conditions have been given in Salibián-Barrera (2000), Salibián-Barrera and Zamar (2002) and Salibián-Barrera *et al.* (2006).

It will become clear from the applications in the next sections that  $\hat{\boldsymbol{\theta}}_n^{R*}$  is not only much easier to compute than  $\hat{\boldsymbol{\theta}}_n^*$ , but also more numerically stable (more robust). Following Salibián-Barrera and Zamar (2002), we will therefore refer to the resulting inference method as the Fast and Robust Bootstrap (FRB).

Section 2 will consider the FRB in the linear regression setting, for which it was originally introduced. We will review the general results for standard errors and confidence intervals of the regression parameters, and we will discuss Salibián-Barrera's (2005) adaptation that can be used to estimate p-values for hypothesis tests in the linear model. Additionally, we consider the potential of the method for obtaining

robust prediction intervals. Section 3 then focuses on applications of the FRB in multivariate methods. First we review how the method was used in the context of robust principal components analysis (PCA) in Salibián-Barrera *et al.* (2006). Next we reconsider its use for inference purposes in robust multivariate regression as described in Van Aelst and Willems (2005). Finally, we propose the application of the bootstrap method as a means of estimating the classification error in robust discriminant analysis. Section 4 illustrates the FRB method on two real-data examples, while Section 5 concludes.

## 2 Robust regression

Let  $\mathcal{Z}_n = \{(y_1, \mathbf{z}_1^t)^t, \dots, (y_n, \mathbf{z}_n^t)^t\}$  be a sample of independent random vectors where  $\mathbf{x}_i = (1, \mathbf{z}_i^t)^t \in \mathbb{R}^p$ .

We will assume that they follow the linear regression model

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}_0 + \sigma_0 \varepsilon_i, \quad i = 1, \dots, n, \quad (7)$$

where the vector of regression coefficients  $\boldsymbol{\beta}_0$  is the parameter of interest. Ideally, one would like to assume that  $y_i$  and  $\mathbf{z}_i$  are independent, with joint distribution function  $H_c(y, \mathbf{z}) = F_c(y) G_c(\mathbf{z})$ , where  $F_c$  is a particular symmetric distribution (typically the standard normal distribution). However, to allow for the occurrence of outliers and other departures from the classical model we will only assume that the distribution  $H$  of the vectors  $(y_i, \mathbf{z}_i^t)^t$  belongs to the contamination neighbourhood

$$\mathcal{H}_\epsilon = \{H = (1 - \epsilon)H_c + \epsilon H^*\} \quad (8)$$

where  $0 \leq \epsilon < 1/2$ , and  $H^*$  is an arbitrary and unspecified distribution.

In the rest of this section we will focus on MM-regression estimators (Yohai 1987). However, all results are valid as well for S-estimators (Rousseeuw and Yohai 1984), since these can be seen as a special case of MM-estimators. In addition, similar results can be derived for other estimators that can

be represented by a smooth fixed-point equation like (1), for example:  $\tau$ -estimators (Yohai and Zamar 1988) or GM-estimates (e.g. Krasker and Welsch 1982).

MM-estimates are based on two loss functions  $\rho_0 : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\rho_1 : \mathbb{R} \rightarrow \mathbb{R}_+$  which determine the breakdown point and the efficiency of the estimate, respectively. More precisely, assume that  $\rho_0$  and  $\rho_1$  satisfy:

(R1)  $\rho$  is symmetric, twice continuously differentiable and  $\rho(0) = 0$ ,

(R2)  $\rho$  is strictly increasing on  $[0, c]$  and constant on  $[c, \infty)$  for some finite constant  $c$ .

The MM-estimate  $\hat{\beta}_n$  satisfies

$$\frac{1}{n} \sum_{i=1}^n \rho_1' \left( \frac{y_i - \mathbf{x}_i^t \hat{\beta}_n}{\hat{\sigma}_n} \right) \mathbf{x}_i = \mathbf{0}, \quad (9)$$

where  $\hat{\sigma}_n$  is a scale S-estimate. That is,  $\hat{\sigma}_n$  minimizes the M-scale  $\hat{\sigma}_n(\beta)$  implicitly defined by the equation

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{y_i - \mathbf{x}_i^t \beta}{\hat{\sigma}_n(\beta)} \right) = b. \quad (10)$$

For future reference, let  $\tilde{\beta}_n$  be the S-regression estimate defined by

$$\tilde{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \hat{\sigma}_n(\beta).$$

It is easy to see that  $\tilde{\beta}_n$  also satisfies

$$\frac{1}{n} \sum_{i=1}^n \rho_0' \left( \frac{y_i - \mathbf{x}_i^t \tilde{\beta}_n}{\hat{\sigma}_n} \right) \mathbf{x}_i = \mathbf{0}, \quad (11)$$

In order for  $\hat{\sigma}_n$  to be consistent under the central model we set  $E_{F_c}[\rho_0(u)] = b$ . The breakdown point of the S-estimator is  $\min(b/\rho_0(\infty), 1 - b/\rho_0(\infty))$ , and thus we normally choose  $\rho_0$  so that  $b = \rho_0(\infty)/2$ .

The MM-estimator  $\hat{\beta}_n$  inherits the breakdown point of the initial S-estimator, while its efficiency at the central model depends only on  $\rho_1$ . In this way these estimators, unlike the S-estimators, are able to achieve simultaneous high-breakdown point and high efficiency at the central model.



Under the assumption that the data follow the central model in (8) the asymptotic distribution of MM-estimates for random and fixed designs has been studied by Yohai (1987) and Salibián-Barrera (2006a), respectively. However this assumption does not hold in typical situations when one wishes to use highly robust MM-estimates. The FRB yields a consistent estimate for the covariance of  $\hat{\beta}_n$  under rather general conditions, including the case  $H \in \mathcal{H}_\epsilon$ .

A widely used family of functions  $\rho$  that satisfy (R1) and (R2) above was proposed by Beaton and Tukey (1974) and is commonly referred to as Tukey's Biweight family:

$$\rho(u) = \begin{cases} 3(u/d)^2 - 3(u/d)^4 + (u/d)^6 & \text{if } |u| \leq d \\ 1 & \text{if } |u| > d, \end{cases} \quad (12)$$

where  $d > 0$  is a fixed constant. An alternative is provided by Yohai and Zamar (1998), whose Optimal weight functions minimize the maximum asymptotic bias of the MM-estimates for a fixed Gaussian efficiency.

### 2.1 Standard errors and confidence intervals

In this section we will consider constructing confidence intervals for the regression coefficients based on MM-regression estimators. We identify two different ways to construct confidence intervals based on MM-estimators: (i) using the fact that MM-estimators are asymptotically normally distributed, we can build a  $(1 - \alpha)\%$  confidence interval for the  $j$ -th regression coefficient  $\beta_{0,j}$  of the form  $\hat{\beta}_{n,j} \pm z_{\alpha/2} \hat{\sigma}_j / \sqrt{n}$ , where  $z_{\alpha/2}$  is the standard normal quantile,  $\hat{\beta}_{n,j}$  is the  $j$ -th coordinate of  $\hat{\beta}_n$ , and  $\hat{\sigma}_j$  is an estimate of its asymptotic standard error; and (ii) we can use the quantiles of the estimated distribution of  $\hat{\beta}_{n,j}$ .

We can use the bootstrap to either obtain  $\hat{\sigma}_j$  or an estimate of the distribution of  $\hat{\beta}_{n,j}$ . Since for these estimators the bootstrap distribution of  $\hat{\beta}_n$  cannot be computed in closed form, we use the computer to

simulate it in the usual way. The rest of this section describes how to use the FRB in this context to solve the problems mentioned in Section 1.

Recall that  $\hat{\beta}_n$  denotes the MM-regression estimate,  $\hat{\sigma}_n$  is the S-scale and  $\tilde{\beta}_n$  the S-regression estimate. Since in general the asymptotic behaviour of  $\hat{\beta}_n$  depends on that of the scale estimator  $\hat{\sigma}_n$  (Carroll 1978, 1979; Carroll and Welsh 1988; Salibián-Barrera 2000; and Croux *et al.* 2003) we need to consider all the equations for  $\hat{\beta}_n$  and  $\hat{\sigma}_n$  when bootstrapping these estimators.

For each pair  $(y_i, \mathbf{x}_i^t)^t$ ,  $i = 1, \dots, n$ , define the residuals associated with  $\hat{\beta}_n$  and  $\tilde{\beta}_n$ :  $r_i = y_i - \mathbf{x}_i^t \hat{\beta}_n$  and  $\tilde{r}_i = y_i - \mathbf{x}_i^t \tilde{\beta}_n$ , respectively. Note that  $\hat{\beta}_n$ ,  $\hat{\sigma}_n$  and  $\tilde{\beta}_n$  can be formally represented as the result of a weighted least squares fit. For  $i = 1, \dots, n$  define the weights  $\omega_i$ ,  $v_i$  and  $\tilde{\omega}_i$  as

$$\begin{aligned}\omega_i &= \rho'_1(r_i/\hat{\sigma}_n)/r_i, \\ v_i &= \frac{\hat{\sigma}_n}{n b} \rho_0(\tilde{r}_i/\hat{\sigma}_n)/\tilde{r}_i, \\ \tilde{\omega}_i &= \rho'_0(\tilde{r}_i/\hat{\sigma}_n)/\tilde{r}_i.\end{aligned}\tag{13}$$

Simple computations yield the following representation of equations (9), (10) and (11):

$$\hat{\beta}_n = \left( \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i y_i, \tag{14}$$

$$\hat{\sigma}_n = \sum_{i=1}^n v_i (y_i - \mathbf{x}_i^t \tilde{\beta}_n), \tag{15}$$

$$\tilde{\beta}_n = \left( \sum_{i=1}^n \tilde{\omega}_i \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \tilde{\omega}_i \mathbf{x}_i y_i. \tag{16}$$

Hence, by setting  $\hat{\theta}_n = (\hat{\beta}_n^t, \hat{\sigma}_n, \tilde{\beta}_n^t)^t$ , MM-estimates can be written in the general form given by (1).

Let  $\mathcal{Z}_n^* = \{(y_1^*, \mathbf{z}_1^{*t})^t, \dots, (y_n^*, \mathbf{z}_n^{*t})^t\}$  be a bootstrap sample from the observations. The corresponding one-step estimates  $\hat{\beta}_n^{1*}$ ,  $\hat{\sigma}_n^{1*}$  and  $\tilde{\beta}_n^{1*}$  as defined by (3) are obtained as follows:

$$\hat{\beta}_n^{1*} = \left( \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \mathbf{x}_i^{*t} \right)^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* y_i^*, \quad (17)$$

$$\hat{\sigma}_n^{1*} = \sum_{i=1}^n v_i^* (y_i^* - \mathbf{x}_i^{*t} \tilde{\beta}_n) \quad (18)$$

$$\tilde{\beta}_n^{1*} = \left( \sum_{i=1}^n \tilde{\omega}_i^* \mathbf{x}_i^* \mathbf{x}_i^{*t} \right)^{-1} \sum_{i=1}^n \tilde{\omega}_i^* \mathbf{x}_i^* y_i^*, \quad (19)$$

where  $\omega_i^* = \rho_1' (r_i^* / \hat{\sigma}_n) / r_i^*$ ,  $v_i^* = \hat{\sigma}_n \rho_0 (\tilde{r}_i^* / \hat{\sigma}_n) / (n b \tilde{r}_i^*)$ ,  $\tilde{\omega}_i^* = \rho_0' (\tilde{r}_i^* / \hat{\sigma}_n) / \tilde{r}_i^*$ ,  $r_i^* = y_i^* - \mathbf{x}_i^{*t} \hat{\beta}_n$ , and  $\tilde{r}_i^* = y_i^* - \mathbf{x}_i^{*t} \tilde{\beta}_n$  for  $1 \leq i \leq n$ . The FRB recalculations for the MM-regression coefficients  $\hat{\beta}_n$ , as defined by (6), are then given by

$$\hat{\beta}_n^{R*} = \hat{\beta}_n + \mathbf{M}_n (\hat{\beta}_n^{1*} - \hat{\beta}_n) + \mathbf{d}_n (\hat{\sigma}_n^{1*} - \hat{\sigma}_n) + \widetilde{\mathbf{M}}_n (\tilde{\beta}_n^{1*} - \tilde{\beta}_n), \quad (20)$$

where

$$\mathbf{M}_n = \hat{\sigma}_n \left( \sum_{i=1}^n \rho_1'' (r_i / \hat{\sigma}_n, \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i^t, \quad \widetilde{\mathbf{M}}_n = 0, \quad (21)$$

$$\mathbf{d}_n = a_n^{-1} \left( \sum_{i=1}^n \rho_1'' (r_i / \hat{\sigma}_n, \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \rho_1'' (r_i / \hat{\sigma}_n, \mathbf{x}_i) r_i \mathbf{x}_i, \quad (22)$$

$$a_n = \frac{1}{n} \frac{1}{b} \sum_{i=1}^n [\rho_0' (\tilde{r}_i / \hat{\sigma}_n) \tilde{r}_i / \hat{\sigma}_n]. \quad (23)$$

Note that since  $\widetilde{\mathbf{M}}_n = 0$ , in practice we do not need to recalculate  $\tilde{\beta}_n$  through (19) when we are only interested in doing inference based on  $\hat{\beta}_n$ .

The empirical standard deviation of the  $\hat{\beta}_{n,j}^{R*}$  values then provides the standard error estimate  $\hat{\sigma}_j$ , whereas their empirical quantiles estimate the quantiles of  $\hat{\beta}_{n,j}$ 's sampling distribution. The number of bootstrap recalculations required for accurate estimates is similar as in classical bootstrap. Typically, a few hundreds suffices for standard error estimates, while at least a thousand is often recommended

for quantile estimates. The number of samples should be increased with the dimension of the problem, however, when interested in simultaneous inference for the different parameters.

Note the computational ease and stability of the FRB estimates in (20):

- *computational cost*: for each bootstrap sample, instead of solving the non-linear system of equations (9) to (11), to compute  $\hat{\beta}_n^{R*}$  we only need to solve the linear system of equations corresponding to (17) and calculate the weighted average (18). The correction factors  $\mathbf{M}_n$ ,  $\mathbf{d}_n$  and  $a_n$  arise from two linear systems and a weighted average respectively and are computed only once with the full sample.
- *numerical stability*: the FRB method is effectively as stable as the estimator itself. If the MM-estimator was able to resist the outlying observations, then the latter are associated with large residuals ( $r_i$  and  $\tilde{r}_i$ ) and thus small or even zero weights in (14)-(16). Indeed, if  $\rho_0$  and  $\rho_1$  satisfy (R1) and (R2), we have

$$\rho'_1(t)/t \xrightarrow{|t| \rightarrow \infty} 0, \quad \rho_0(t)/t \xrightarrow{|t| \rightarrow \infty} 0 \quad \text{and} \quad \rho'_0(t)/t \xrightarrow{|t| \rightarrow \infty} 0.$$

The outliers take along their small weights into the one-step bootstrap approximations of (17)-(19). Hence, the influence of harmful outlying observations is limited, regardless of the number in which they appear in the bootstrap sample. The linear correction factors, finally, do not depend on the bootstrap sample and will not break down if the original MM-estimates did not break down.

The analysis above on computational ease and numerical stability can similarly be applied to other applications of the FRB, such as the multivariate models discussed in Section 3.

*2.1.1 Consistency of the fast and robust bootstrap* The next theorem (proved in Salibián-Barrera and Zamar (2002)) shows that, under certain regularity conditions, if the S- and MM-estimators are consistent, then the asymptotic distribution of the FRB coincides with that of the MM-regression estimator. Sufficient conditions for the S-estimator to be consistent and asymptotically normally distributed are

given in Davies (1990), while MM-estimators were studied by Yohai (1987), Salibian-Barrera (2000) and Salibian-Barrera (2006a). See also Omelka and Salibian-Barrera (2006) for verifiable regularity conditions to obtain consistency and asymptotic normality for S- and MM-regression estimators uniformly over sets of distributions of the form (8).

**Theorem 1** (Salibian-Barrera and Zamar 2002) - *Let  $\rho_0$  and  $\rho_1$  satisfy (R1) and (R2) with continuous third derivatives. Let  $\hat{\beta}_n$  be the MM-regression estimator,  $\hat{\sigma}_n$  the S-scale and  $\tilde{\beta}_n$  the associated S-regression estimator and assume that  $\hat{\beta}_n \xrightarrow{P} \beta$ ,  $\hat{\sigma}_n \xrightarrow{P} \sigma$  and  $\tilde{\beta}_n \xrightarrow{P} \tilde{\beta}$ . Then, under certain regularity conditions, along almost all sample sequences,  $\sqrt{n}(\hat{\beta}_n^{R*} - \hat{\beta}_n)$  converges weakly, as  $n$  goes to infinity, to the same limit distribution as  $\sqrt{n}(\hat{\beta}_n - \beta)$ .*

**2.1.2 Robustness of confidence intervals** Confidence intervals for the regression parameters can be based on the quantiles of the bootstrap distribution estimate. It is clear from the discussion above that the FRB method will yield estimates of the distribution quantiles that are more reliable (less affected by outliers) than those of the classical bootstrap. In this section we give a formal result for the breakdown point of the quantile estimates of the FRB.

Following Singh (1998) we define the upper breakdown point of a quantile estimate  $\hat{q}_t$  as the minimum proportion of asymmetric contamination that can drive it over any finite bound. There are two closely related scenarios in which the bootstrap quantile estimates can break down. The first unfavourable situation is when the proportion of outliers in the original data is larger than the breakdown point of the estimate. In this case the estimate may already be unreliable, and so can be the inferences derived from it. The second case is related to the number of outliers appearing in the bootstrap samples. If we let  $\tau^*$  denote the expected proportion of bootstrap samples that contain more outliers than the breakdown point of the estimate, then the estimate  $\hat{q}_t$  may be severely affected by the outliers when  $\tau^* > t$ . The

next theorem shows that the FRB is not affected as severely as the classical bootstrap by the second scenario.

Note that the breakdown point of the FRB quantile estimates can be affected by the geometric characteristics of a sample in the same way as the breakdown point of robust regression estimates. Recall that  $k$  points in  $\mathbb{R}^p$  are said to be in *general position* if no subset of size  $p + 1$  of them determines an affine subspace of dimension  $p$  (Rousseeuw and Leroy 1987). The breakdown point of the quantile estimates based on the FRB when the sample is in general position are given by the following theorem.

**Theorem 2** (Salibian-Barrera and Zamar 2002) *Assume that the sample  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is in general position, and let  $\hat{\beta}_n$  be an MM-regression estimate with breakdown point  $\epsilon^*$ . The breakdown point of the FRB estimate for the  $t$ -th quantile of the sampling distribution of  $\hat{\beta}_{nj}$ ,  $j = 1, \dots, p$  is given by  $\min(\epsilon^*, \epsilon_{FRB}^t)$ , where  $\epsilon_{FRB}^t$  satisfies*

$$\epsilon_{FRB}^t = \inf \{ \delta \in [0, 1] : P[\text{Binomial}(n, \delta) \geq n - p] \geq t \} . \quad (24)$$

Singh (1998) showed that the upper breakdown point  $\epsilon_C^t$  of the bootstrap estimate  $\hat{q}_t$  of  $q_t$  satisfies

$$\epsilon_C^t = \inf \{ \delta \in [0, 1] : P[\text{Binomial}(n, \delta) \geq \lceil \epsilon^* n \rceil] \geq t \} , \quad (25)$$

where  $\lceil x \rceil$  denotes the smallest integer larger or equal to  $x$ . Since  $\lceil \epsilon^* n \rceil \leq \lfloor n/2 \rfloor < n - p$  for  $n > 2p$ , we immediately see from (24) and (25) that  $\epsilon_C^t < \epsilon_{FRB}^t$  and thus that the FRB quantile estimators are more robust than those based on the classical bootstrap. A comparison of  $\epsilon_C^t$  and  $\epsilon_{FRB}^t$  for different values of  $t$ , sample sizes ( $n$ ) and number of covariates ( $p$ ) can be found in Table 1.

## 2.2 Hypothesis tests

In this section we show how to use the FRB to estimate the distribution of robust test statistics for the linear regression model. Following Salibian-Barrera (2005) we will focus on scores-type tests (Markatou

**Table 1** Comparison of quantile upper breakdown points for MM-regression estimates with 50% breakdown point and 95% efficiency at the central model.

$p$	$n$	FR Bootstrap			Classical Bootstrap		
		$\hat{q}_{0.005}$	$\hat{q}_{0.025}$	$\hat{q}_{0.05}$	$\hat{q}_{0.005}$	$\hat{q}_{0.025}$	$\hat{q}_{0.05}$
2	10	0.456	0.500	0.500	0.128	0.187	0.222
	20	0.500	0.500	0.500	0.217	0.272	0.302
	30	0.500	0.500	0.500	0.265	0.313	0.339
5	10	0.191	0.262	0.304	0.011	0.025	0.036
	20	0.500	0.500	0.500	0.114	0.154	0.177
	30	0.500	0.500	0.500	0.185	0.226	0.249
	100	0.500	0.500	0.500	0.368	0.398	0.414
10	20	0.257	0.315	0.347	0.005	0.012	0.018
	50	0.500	0.500	0.500	0.180	0.212	0.230
	100	0.500	0.500	0.500	0.294	0.322	0.336

*et al.* 1991), although the same approach can be used with other tests (e.g. robust dispersion and Wald type tests). The finite-sample distribution of these test statistics is generally unknown and their p-values are approximated using their asymptotic distribution. As in the case of linear regression estimators, the asymptotic distribution of robust tests has been studied under the restrictive assumption that the scale of the errors is known or that it can be estimated without affecting the asymptotic behavior of the test statistic. This is generally only true when the errors have a symmetric distribution. In the rest of this section we describe how to use the FRB to estimate the p-values of scores-type tests for the linear regression model (7). Partition the vector of parameters  $\beta_0 = (\beta_{0,1}^t, \beta_{0,2}^t)^t$  where  $\beta_{0,2} \in \mathbb{R}^q$ ,  $q < p$ .

Without loss of generality assume that the covariates have been re-arranged so that the hypotheses to be tested are

$$H_0 : \beta_{0,2} = \mathbf{0} \quad \text{versus} \quad H_a : \beta_{0,2} \neq \mathbf{0}. \quad (26)$$

In a similar way, write  $\mathbf{x}_i = (\mathbf{x}_{i(1)}^t, \mathbf{x}_{i(2)}^t)^t$  where  $\mathbf{x}_{i(2)} \in \mathbb{R}^q$ . Let  $\hat{\beta}_n^{(0)} \in \mathbb{R}^{p-q}$  and  $\hat{\beta}_n^{(a)} \in \mathbb{R}^p$  be the robust estimates under  $H_0$  and  $H_a$ . Correspondingly, let  $r_i^{(0)}$  and  $r_i^{(a)}$  be the associated residuals:

$$r_i^{(a)} = y_i - \mathbf{x}_i^t \hat{\beta}_n^{(a)}, \quad r_i^{(0)} = y_i - \mathbf{x}_{i(1)}^t \hat{\beta}_n^{(0)},$$

for  $i = 1, \dots, n$ . Also, let  $\hat{\sigma}_n^{(a)}$  be the scale estimate calculated under  $H_a$ . Scores-type tests (Markatou *et al.* 1991) are defined as follows:

$$W_n^2 = n^{-1} [\boldsymbol{\Sigma}_n^{(0)}]^t \hat{\mathbf{U}}^{-1} [\boldsymbol{\Sigma}_n^{(0)}],$$

with

$$\boldsymbol{\Sigma}_n^{(0)} = \sum_{i=1}^n \rho_1'((y_i - \mathbf{x}_{i(1)}^t \hat{\beta}_n^{(0)}) / \hat{\sigma}_n^{(a)}) \mathbf{x}_{i(2)},$$

and

$$\hat{\mathbf{U}} = \hat{\mathbf{Q}}_{22} - \hat{\mathbf{M}}_{21} \hat{\mathbf{M}}_{11}^{-1} \hat{\mathbf{Q}}_{12} - \hat{\mathbf{Q}}_{21} \hat{\mathbf{M}}_{11}^{-1} \hat{\mathbf{M}}_{12} + \hat{\mathbf{M}}_{21} \hat{\mathbf{M}}_{11}^{-1} \hat{\mathbf{Q}}_{11} \hat{\mathbf{M}}_{11}^{-1} \hat{\mathbf{M}}_{12},$$

where  $\hat{\mathbf{M}}$  and  $\hat{\mathbf{Q}}$  are the empirical estimates of

$$\mathbf{M} = E[\rho_1''(r) \mathbf{x} \mathbf{x}^t] = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix},$$

and

$$\mathbf{Q} = E[\rho_1'(r)^2 \mathbf{x} \mathbf{x}^t] = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix},$$

respectively.

Under  $H_0$ ,  $\hat{\beta}_n^{(0)}$  is expected to approximately solve the score equations (even the part involving the  $\mathbf{x}_{i(2)}$ 's). Hence, in this case  $W_n^2$  will tend to be small and large values of  $W_n^2$  indicate evidence against  $H_0$ .



When the errors have a symmetric distribution around zero,  $W_n^2$  has, under  $H_0$ , an asymptotic  $\chi^2$  distribution with  $q$  degrees of freedom (Markatou *et al.* 1991). It is easy to see (Salibian-Barrera 2005) that when the errors are not symmetrically distributed (due to the presence of outliers or other forms of contamination)  $W_n^2$  may have a different asymptotic distribution. We propose to use the FRB to estimate this distribution.

When bootstrapping tests it is important that the bootstrap re-sampling scheme resembles the null hypothesis  $H_0$  (even when the data actually follow the alternative hypothesis  $H_a$ , see Fisher and Hall (1990) and Hall and Wilson (1991)). In other words, to properly simulate the null distribution of the test statistics we have to draw bootstrap samples from data that follow  $H_0$ . One way to do this in the context of non-parametric bootstrap for linear regression is to create a data set with this property by adding the residuals obtained under  $H_a$  to the fitted values obtained under  $H_0$ . Furthermore, to construct bootstrap samples following  $H_0$  we add a bootstrap sample of the residuals obtained under  $H_a$  to the fitted values computed under  $H_0$ . These bootstrapped data will follow the null model when the original sample follows either  $H_0$  or  $H_a$ .

Since the robust scores-type statistics are of the form

$$W_n^2 = W_n^2(y_1, \dots, y_n, \mathbf{x}_1, \dots, \mathbf{x}_n, \hat{\beta}_n^{(0)}, \hat{\beta}_n^{(a)})$$

we need to estimate the null distribution of  $(y_i, \mathbf{x}_i)$ ,  $\hat{\beta}_n^{(0)}$  and  $\hat{\beta}_n^{(a)}$ . The following “null data” follow  $H_0$ :

$$\tilde{y}_i = \mathbf{x}_i^t \hat{\beta}_n^{(0)} + r_i^{(a)}, \quad (27)$$

for  $i = 1, \dots, n$ . Note that these “null data” satisfy  $H_0$  approximately by construction even when the original data  $(y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$  follow  $H_a$ .

In order to use the FRB to re-calculate the  $\hat{\beta}_n$ ’s we first calculate the robust estimates under the restricted and full models with the complete “null” data set, that is with  $(\tilde{y}_i, \mathbf{x}_i)$ , where  $\tilde{y}_i$  is given by (27).

Let  $\ddot{\beta}_n^{(0)}$  and  $\ddot{\beta}_n^{(a)}$  be the MM-regression estimates calculated under  $H_0$  and  $H_a$  respectively. Similarly,  $\ddot{\sigma}_n^{(a)}$  denotes the scale estimate calculated under the full model on the “null data”. Let  $r_1^{*(a)}, \dots, r_n^{*(a)}$  be a bootstrap sample from the residuals  $r_1^{(a)}, \dots, r_n^{(a)}$  and construct the bootstrap observations

$$\tilde{y}_i^* = \mathbf{x}_i^t \hat{\beta}_n^{(0)} + r_i^{*(a)}, \quad i = 1, \dots, n.$$

We can apply the FRB to  $\ddot{\beta}_n^{(0)}$ ,  $\ddot{\beta}_n^{(a)}$  and  $\ddot{\sigma}_n^{(a)}$  to obtain fast and robust bootstrap recalculations and evaluate the test statistics with them. These values will be the bootstrap estimate of the sampling distribution of the test statistics. The consistency of the FRB under this re-sampling scheme (which is equivalent to bootstrapping linear regression estimators with fixed designs (Freedman 1981)) was studied in Salibián-Barrera (2006b).

Let  $(\tilde{y}_i^*, \mathbf{x}_i)$  be a bootstrap sample where  $\tilde{y}_i^* = \mathbf{x}_i^t \hat{\beta}_n^{(0)} + r_i^{*(a)}$ . Let  $\ddot{\beta}_n^{R*(0)}$  and  $\ddot{\beta}_n^{R*(a)}$  be the FRB re-calculations under  $H_0$  and  $H_a$  respectively. The FRB re-calculated test statistic is

$$W_n^{2R*} = n^{-1} \boldsymbol{\Sigma}_n^{R*} (\ddot{\beta}_n^{R*(0)})^t \mathbf{U}^{R*-1} \boldsymbol{\Sigma}_n^{R*} (\ddot{\beta}_n^{R*(0)}),$$

with  $\mathbf{U}^{R*}$  the corresponding matrix, and

$$\boldsymbol{\Sigma}_n^{R*}(\beta) = \sum_{i=1}^n \rho_1'((\tilde{y}_i^* - \beta^t \mathbf{x}_i) / \ddot{\sigma}_n^{(a)}) \mathbf{x}_i \mathbf{x}_i^t.$$

The FRB p-value estimate is

$$\hat{p} = \# \{W_n^{2R*} > W_n^2\} / B, \quad (28)$$

where  $B$  is the number of bootstrap samples used. In Salibián-Barrera (2005) it is shown that these bootstrap p-values maintain the nominal level even under the presence of contamination, and are equivalent to those obtained with the asymptotic  $\chi^2$  approximation when the data do not contain outliers.

### 2.3 Mean response and prediction intervals

The bootstrap distribution obtained from the FRB estimates (20) can be used to construct confidence intervals for the mean response at a fixed  $\mathbf{x}_0$  in the predictor space. Similarly as for confidence intervals of the regression parameters, there are two ways to construct such intervals based on the FRB estimates: (i) using the asymptotic normality of the MM-estimators, we can build a  $(1 - \alpha)\%$  confidence interval for  $\mu_0 = \mathbf{x}_0^t \boldsymbol{\beta}_0$  of the form  $\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n \pm z_{\alpha/2} \sqrt{\mathbf{x}_0^t \hat{\Sigma} \mathbf{x}_0}$ , where  $\hat{\Sigma}$  is a bootstrap estimate of the covariance matrix of  $\hat{\boldsymbol{\beta}}_n$ ; and (ii) we can use the bootstrap to estimate the distribution of  $\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n$  and use quantiles of this distribution estimate to construct confidence intervals. The consistency of the FRB for the MM-estimator of the regression parameters in Theorem 1 immediately implies consistency of the FRB at  $\mathbf{x}_0$  under the same conditions, in the sense that the distribution  $\sqrt{n}(\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n^{R*} - \mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n)$  converges weakly to the same limit distribution as  $\sqrt{n}(\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n - \mu_0)$ . Moreover, for samples in general position the FRB estimates for the quantiles of the sampling distribution of  $\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n$  will have the breakdown point given by Theorem 2. A prediction interval for the unobserved response  $y_0$  at  $\mathbf{x}_0$  can be constructed as  $\mathbf{x}_0^t \hat{\boldsymbol{\beta}}_n \pm z_{\alpha/2} \sqrt{\mathbf{x}_0^t \hat{\Sigma} \mathbf{x}_0 + \hat{\sigma}_n}$  where  $\hat{\sigma}_n$  is the scale S-estimate as before.

To investigate the performance of these confidence and prediction intervals we performed a short simulation study. We generated 5000 datasets of size  $n = 40, 100$  and dimension  $p = 2, 5$  (including the intercept). The explanatory variables were generated from the standard normal distribution  $N(0, 1)$ . In the first simulation setting, the errors were generated according to  $(n - 2m) N(0, 1) + m N(-4, 0.1^2) + m N(4, 0.1^2)$  which means that  $(n - 2m)$  points are drawn from  $N(0, 1)$  and  $m$  points from each of the two contaminating distributions, leading to symmetrically distributed outliers. In the second simulation setting, the errors were generated according to  $(n - 2m) N(0, 1) + 2m N(4, 0.1^2)$  which produces asymmetric outliers. For both settings we considered  $m = 0, .05n$  and  $.10n$  corresponding to 0%, 10%, and 20% of contamination. 95% and 99% confidence and prediction intervals were constructed at  $\mathbf{x}_0 = (1, \mathbf{d})^t$

**Table 2** Coverage and length of 99% FRB based confidence and prediction intervals at  $\mathbf{x}_0 = (1, \mathbf{0})^t =$  and  $\mathbf{x}_0 = (1, \mathbf{3})^t$ . Samples of size  $n = 100$  with  $2m$  symmetrically distributed outliers in  $p = 5$  dimensions for  $m = 0, 5$  and  $10$ .

$\mathbf{x}_0$	$m$	FRB-quantile CI	FRB-SE CI	FRB Pred Int
$(1, \mathbf{0})^t$	0	0.985 (0.547)	0.986 (0.547)	0.987 (5.051)
	5	0.992 (0.694)	0.993 (0.691)	0.994 (5.952)
	10	0.999 (1.037)	0.999 (1.032)	0.999 (7.240)
$(1, \mathbf{3})^t$	0	0.987 (3.399)	0.985 (3.357)	0.985 (6.056)
	5	0.990 (4.395)	0.987 (4.331)	0.993 (7.359)
	10	0.991 (6.580)	0.994 (6.469)	0.998 (9.721)

with  $\mathbf{d} = (d, \dots, d) \in \mathbb{R}^{p-1}$  for  $d = 0, 1, 2, 3$ . We used MM-estimators with  $\rho$  functions in Tukey's Bi-weight family (12), tuned to have 50% breakdown point and 95% Gaussian efficiency. Since for  $n = 100$  and  $p = 5$ , 20% of outliers in the original sample do not influence the classical bootstrap distribution estimator noticeably (e.g. the 0.5% quantile does not breakdown, as can be seen from Table 1), there will not yet be problems with numerical stability of the classical bootstrap, and hence the classical bootstrap here shows similar behavior as the FRB but needs much more time.

For symmetrically distributed outliers, Table 2 shows the coverage level and average length of 99% confidence and prediction intervals at  $\mathbf{x}_0 = (1, \mathbf{0})^t =$  and  $\mathbf{x}_0 = (1, \mathbf{3})^t$  which are points in the center and at the border of the predictor space. The results shown are for  $n = 100$  and  $p = 5$ . From this table we see that all intervals become more conservative when the amount of contamination increases. The coverage level is comparable throughout the predictor space but not surprisingly the length of the intervals increases with the distance from the center of the predictor space. The increase is more pronounced for the confidence intervals than for the prediction intervals. With this simulation design, the

**Table 3** Coverage and length of 99% FRB based confidence and prediction intervals at  $\mathbf{x}_0 = (1, \mathbf{0})^t$  and  $\mathbf{x}_0 = (1, \mathbf{3})^t$ . Samples of size  $n = 100$  with  $2m$  asymmetric outliers in  $p = 5$  dimensions for  $m = 0, 5$  and  $10$ .

$\mathbf{x}_0$	$m$	FRB-quantile CI	FRB-SE CI	FRB Pred Int
$(1, \mathbf{0})^t$	0	0.985 (0.547)	0.986 (0.547)	0.987 (5.051)
	5	0.970 (0.726)	0.959 (0.723)	0.996 (5.965)
	10	0.373 (0.959)	0.327 (0.952)	0.997 (7.230)
$(1, \mathbf{3})^t$	0	0.987 (3.399)	0.985 (3.357)	0.985 (6.056)
	5	0.988 (4.470)	0.987 (4.411)	0.994 (7.414)
	10	0.986 (6.670)	0.985 (6.570)	0.997 (9.780)

confidence intervals based on bootstrap quantile estimates tend to be slightly larger than the confidence intervals based on a bootstrap estimate of the standard error.

Table 3 shows the results for asymmetric outliers. The coverage of the confidence intervals at  $\mathbf{x}_0 = (1, \mathbf{3})^t$  is a little bit below the nominal level of 99% but the performance is still acceptable. However, for  $\mathbf{x}_0 = (1, \mathbf{0})^t$  the coverage is very low. This negative effect is due to the large bias on the MM-estimate of the intercept caused by the outliers in this simulation setting. The confidence and prediction intervals based on FRB only reflect sampling variability of the estimates but do not take bias into account. Therefore, if asymmetrically distributed outliers may be present in the data, we recommend that the FRB be combined with globally robust inference methods as those discussed in Adrover et al. (2004) which are designed to reflect both sampling variability and bias.

### 3 Robust multivariate methods

Although not as widespread as in linear regression, the bootstrap has various useful applications in multivariate methods. In this section, we will discuss three of these: (1) inference and stability assess-

ment in principal components analysis; (2) inference in multivariate linear regression; and (3) error rate estimation in discriminant analysis.

Applying classical bootstrap to robust multivariate estimators comes with the same problems as in robust regression: a high computational cost and low breakdown. The FRB method offers a much more practical and reliable alternative. It can be applied in the case of some well-known robust multivariate estimators, most notably multivariate S-estimators (Rousseeuw and Leroy 1987; Davies 1987) or multivariate MM-estimators (Tatsuoka and Tyler 2000). Similarly to the linear regression setting, MM-estimators avoid the S-estimators' trade-off between Gaussian efficiency and breakdown point. However, as the dimension  $p$  of the data increases, the relative efficiency of a high breakdown S-estimator increases as well (see e.g. Salibián-Barrera *et al.*, 2006). Hence, the trade-off is less severe than in univariate regression. For this reason, and because of their popularity, we will focus on S-estimators in this section. Note however that the FRB can equally well be applied to multivariate MM-estimators (as in Salibián-Barrera *et al.*, 2006), or to other “smooth” estimators for that matter (for example Lopuhaä's (1992) multivariate  $\tau$ -estimators).

### 3.1 Principal components analysis

Principal components analysis is a very common technique in multivariate statistics. It aims to explain the covariance structure of the data by projecting the observations onto a small number of principal components, which are linear combinations of the original variables. Classical PCA is based on the eigenvectors and eigenvalues of the sample covariance or shape matrix. Robust PCA methods which replace the latter by a robust estimate of scatter or shape have been proposed and investigated by many authors starting with Maronna (1976), Campbell (1980) and Devlin *et al.* (1981). More recently Croux

and Haesbroeck (2000) compared several robust estimators in a more formal manner, and suggested the use of S-estimators.

Suppose we have a sample  $\mathcal{Z}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$ , and let  $\rho$  be a function that satisfies (R1) and (R2). Then the S-estimates of location and scatter,  $\hat{\boldsymbol{\mu}}_n$  and  $\hat{\Sigma}_n$ , minimize  $|C|$  subject to

$$\frac{1}{n} \sum_{i=1}^n \rho \left( [(\mathbf{x}_i - T)^t C^{-1} (\mathbf{x}_i - T)]^{\frac{1}{2}} \right) = b \quad (29)$$

among all  $T \in \mathbb{R}^p$  and  $C \in \text{PDS}(p)$ . Here,  $\text{PDS}(p)$  denotes the set of positive definite symmetric  $p \times p$  matrices. The constant  $b$  is usually chosen such that  $b = E_{\Phi}[\rho(\|\mathbf{x}\|)]$ , which ensures consistency at the normal model. The asymptotic breakdown point of  $\hat{\boldsymbol{\mu}}_n$  and  $\hat{\Sigma}_n$  equals  $\min(b/\rho(\infty), 1 - b/\rho(\infty))$ . For any given dimension  $p$ ,  $\rho$  can be tuned in order to achieve a 50% breakdown point.

The corresponding robust PCA method is based on the eigenvalues and eigenvectors of  $\hat{\Sigma}_n$ . Regarding inference, results based on asymptotic normality can be used to construct confidence intervals or to estimate standard errors of the eigenvalues and eigenvectors (see e.g. Croux and Haesbroeck, 2000). However, these results only hold under the assumption of some underlying elliptical distribution, which is of course not appropriate in those cases where robust estimation is most recommended. Inference based on the asymptotic variances derived at the central model may still yield reasonable results for small amounts of contamination when the sample size is large. The bootstrap, however, is expected to work better for smaller sample sizes and for larger deviations from the central model. Moreover, since the bootstrap can estimate the sampling distribution of the estimator of interest, it offers a wider range of inference applications and allows to assess the stability of the PCA results.

Salibian-Barrera *et al.* (2006) used the fast and robust bootstrap for robust PCA based on multivariate MM-estimates. Their discussion and theoretical results are valid for S-estimates as well, since the latter can be seen as a special case of MM-estimates. The FRB method basically generates many

recalculations of  $\widehat{\Sigma}_n$ , and inference or stability assessment is based on the eigenvalues and eigenvectors of these recomputed matrices.

We can fit the multivariate S-estimates into the fixed-point setting of (1) when we take  $\hat{\boldsymbol{\theta}}_n = ((\hat{\boldsymbol{\mu}}_n)^t, \text{vec}(\widehat{\Sigma}_n)^t)^t$ . Indeed, analogously to the regression equations (14)-(16), it can be shown that the estimates  $\hat{\boldsymbol{\theta}}_n$  satisfy the following weighted mean and covariance expressions:

$$\hat{\boldsymbol{\mu}}_n = \left( \sum_{i=1}^n \frac{\rho'(d_i)}{d_i} \right)^{-1} \sum_{i=1}^n \frac{\rho'(d_i)}{d_i} \mathbf{x}_i \quad (30)$$

$$\widehat{\Sigma}_n = \frac{1}{nb} \left( \sum_{i=1}^n p \frac{\rho'(d_i)}{d_i} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)^t + \left( \sum_{i=1}^n s_i \right) \widehat{\Sigma}_n \right) \quad (31)$$

where  $d_i = [(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)^t \widehat{\Sigma}_n^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)]^{1/2}$  and  $s_i = \rho(d_i) - \rho'(d_i)d_i$ .

The one-step bootstrap estimates  $\hat{\boldsymbol{\theta}}_n^{1*}$  are again obtained by taking the bootstrap versions of (30)-(31), but letting each observation keep its original weight. Subsequently we compute (6) to obtain  $\hat{\boldsymbol{\theta}}_n^{R*}$ , and in particular the statistic of interest here,  $\widehat{\Sigma}_n^{R*}$ . Herein, we need the partial derivatives of the function  $\mathbf{g}_n$  as defined by (30)-(31), which are obtained through some tedious though straightforward calculations (see also Van Aelst and Willems, 2004).

Note that, because of the linear correction in (6), the recalculated shape estimates  $\widehat{\Sigma}_n^{R*}$  may not be positive definite and hence the bootstrapped eigenvalue estimates can be negative. In practice this seems to occur very rarely, but care must be taken, especially for small sample sizes. A simple solution is to discard those bootstrap samples where this happens. Alternatively, one can consider transformations of  $\widehat{\Sigma}_n^{R*}$  such as those described in Rousseeuw and Molenberghs (1993).

Salibián-Barrera *et al.* (2006) proved the consistency of the FRB for multivariate MM-estimates of location and scatter. We state their theorem here for the special case of S-estimates. Similarly to the conditions of Theorem 1, it is required that  $\hat{\boldsymbol{\mu}}_n$  and  $\widehat{\Sigma}_n$  be consistent estimates. See Lopuhaä (1989) for weak assumptions under which S-estimates are consistent.



**Theorem 3** (Salibian-Barrera *et al.* 2006) *Let  $\rho$  satisfy (R1) and (R2) and let  $(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n)$  be the corresponding multivariate  $S$ -estimates. Assume that  $\hat{\boldsymbol{\mu}}_n \xrightarrow{P} \boldsymbol{\mu}$ ,  $\hat{\Sigma}_n \xrightarrow{P} \Sigma$ . Then, under certain regularity conditions, along almost all sample sequences the distributions of  $\sqrt{n}(\hat{\boldsymbol{\mu}}_n^{R*} - \hat{\boldsymbol{\mu}}_n)$  and  $\sqrt{n}(\hat{\Sigma}_n^{R*} - \hat{\Sigma}_n)$  converge weakly to the same limit distributions as those of  $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu})$  and  $\sqrt{n}(\hat{\Sigma}_n - \Sigma)$  respectively.*

Bickel and Freedman (1981) show that the bootstrap commutes with smooth functions. This can be used to prove the consistency of bootstrapping the eigenvalues and eigenvectors of  $\hat{\Sigma}_n$ . We do need a restriction to simple eigenvalues though, since otherwise we do not have the necessary smoothness conditions.

**Theorem 4** (Salibian-Barrera *et al.* 2006) *Let  $\lambda_\Sigma$  be a simple eigenvalue of  $\Sigma$ , with normalized eigenvector  $v_\Sigma$ . Then in some neighbourhood  $N(\Sigma)$  of  $\Sigma$  there exists a real-valued function  $\lambda$  and vector function  $v$  such that  $\lambda(\Sigma) = \lambda_\Sigma$  and  $v(\Sigma) = v_\Sigma$ , as well as  $Cv = \lambda v, v^t v = 1$  for all  $C \in N(\Sigma)$ . Assuming the same regularity conditions as in the previous theorem, we have that the distributions of  $\sqrt{n}(\lambda(\hat{\Sigma}_n^{R*}) - \lambda(\hat{\Sigma}_n))$  and  $\sqrt{n}(v(\hat{\Sigma}_n^{R*}) - v(\hat{\Sigma}_n))$  converge weakly to the same limit distributions as those of  $\sqrt{n}(\lambda(\hat{\Sigma}_n) - \lambda(\Sigma))$  and  $\sqrt{n}(v(\hat{\Sigma}_n) - v(\Sigma))$  respectively.*

Theorem 4 serves as the theoretical justification for applying the FRB method in robust PCA.

Examples of (classical) bootstrap applied to (classical) PCA can be found in Diaconis and Efron (1983), Daudin *et al.* (1988) or Beran and Srivastava (1985) amongst others. Salibian-Barrera *et al.* (2006) considered three different bootstrap applications:

1. Assessing the variability of the eigenvalues of  $\hat{\Sigma}_n$ , as estimates of the eigenvalues of the population scatter matrix  $\Sigma$ ; simulations showed that confidence intervals based on the FRB are reasonably short and have good coverage, generally comparable to confidence intervals based on the asymptotic normality approach; both bootstrap and asymptotic inference for eigenvalues is less reliable, however, when the difference between consecutive eigenvalues of  $\hat{\Sigma}_n$  is small.

2. Estimating the distribution of the angles between the eigenvectors of  $\widehat{\Sigma}_n$  and those of  $\Sigma$ ; simulations showed that this is quite efficiently estimated by the distribution of the angles between the FRB recalculated eigenvectors of  $\widehat{\Sigma}_n^{R*}$  and the eigenvectors of  $\widehat{\Sigma}_n$ ; for example, an eigenvector of  $\widehat{\Sigma}_n$  which is relatively aligned with its recalculations based on  $\widehat{\Sigma}_n^{R*}$ , can be considered an accurate estimate of the corresponding eigenvector of  $\Sigma$ .
3. Assessing the variability of  $\widehat{p}_k = (\sum_{i=1}^k \widehat{\lambda}_i) / (\sum_{i=1}^p \widehat{\lambda}_i)$ , where  $\widehat{\lambda}_i; i = 1, \dots, p$  are the ordered eigenvalues of  $\widehat{\Sigma}_n$ ; the statistic  $\widehat{p}_k$  estimates the percentage of variance explained by the first  $k$  robust principal components ( $k = 1, \dots, p-1$ ) and is often used to decide on how many principal components to retain for further analysis; confidence limits accompanying  $\widehat{p}_k$  can give additional information on which to base such a decision; simulations showed that confidence intervals based on the FRB usually outperform the intervals based on asymptotic normality.

Note that Salibián-Barrera *et al.* (2006) in fact worked with the eigenvalues and eigenvectors of the shape component of  $\widehat{\Sigma}_n$  (which is defined as  $\widehat{\Gamma}_n := |\widehat{\Sigma}_n|^{-1/p} \widehat{\Sigma}_n$ ), instead of with those of  $\widehat{\Sigma}_n$  itself. This approach in principle yields exactly the same robust PCA method, although inference for eigenvalues tends to be less troubled by outlier-induced bias.

### 3.2 Multivariate linear regression

Consider now a sample of the form  $\mathcal{Z}_n = \{(\mathbf{y}_1^t, \mathbf{x}_1^t)^t, \dots, (\mathbf{y}_n^t, \mathbf{x}_n^t)^t\} \subset \mathbb{R}^{q+p}$ . The multivariate linear regression model is given by

$$\mathbf{y}_i = \mathcal{B}_0^t \mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (32)$$

where  $\mathcal{B}_0$  is the  $p \times q$  matrix of regression coefficients. It is again assumed that the distribution of the observations belongs to a contamination neighbourhood where the central model is such that the

$q$ -variate error vectors  $\epsilon_i$  have an elliptically contoured density with zero center and scatter parameter  $\Sigma_0 \in \text{PDS}(q)$ .

Quite similar to the univariate situation, least squares estimation of  $\mathcal{B}_0$  is very sensitive to outliers and several robustifications have been proposed. One of these is Van Aelst and Willems' (2005) proposal to incorporate the regression model into the multivariate S-estimates as defined in (29). The resulting multivariate regression S-estimates  $\hat{\mathcal{B}}_n$  and  $\hat{\Sigma}_n$  minimize  $|C|$  subject to

$$\frac{1}{n} \sum_{i=1}^n \rho \left( [(\mathbf{y}_i - B^t \mathbf{x}_i)^t C^{-1} (\mathbf{y}_i - B^t \mathbf{x}_i)]^{\frac{1}{2}} \right) = b \quad (33)$$

among all  $B \in \mathbb{R}^{p \times q}$  and  $C \in \text{PDS}(q)$ .

Robust multivariate regression based on S-estimates has good robustness and efficiency properties (Van Aelst and Willems 2005) and of course allows the application of the FRB method for inference purposes. The estimating equations (1), with  $\hat{\boldsymbol{\theta}}_n = (\text{vec}(\hat{\mathcal{B}}_n)^t, \text{vec}(\hat{\Sigma}_n)^t)^t$ , are now given by

$$\hat{\mathcal{B}}_n = \left( \sum_{i=1}^n \frac{\rho'(d_i)}{d_i} \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n \frac{\rho'(d_i)}{d_i} \mathbf{x}_i \mathbf{y}_i^t \quad (34)$$

$$\hat{\Sigma}_n = \frac{1}{nb} \left( \sum_{i=1}^n p \frac{\rho'(d_i)}{d_i} (\mathbf{y}_i - \hat{\mathcal{B}}_n^t \mathbf{x}_i)(\mathbf{y}_i - \hat{\mathcal{B}}_n^t \mathbf{x}_i)^t + \left( \sum_{i=1}^n s_i \right) \hat{\Sigma}_n \right) \quad (35)$$

where we redefined  $d_i$  as  $d_i = [(\mathbf{y}_i - \hat{\mathcal{B}}_n^t \mathbf{x}_i)^t \hat{\Sigma}_n^{-1} (\mathbf{y}_i - \hat{\mathcal{B}}_n^t \mathbf{x}_i)]^{1/2}$  and where  $s_i = \rho(d_i) - \rho'(d_i)d_i$  as before.

The FRB recalculations  $\hat{\mathcal{B}}_n^{R*}$  and  $\hat{\Sigma}_n^{R*}$  are again defined by (6). Expressions for the partial derivatives of the function  $\mathbf{g}_n$  for this particular application are given in Van Aelst and Willems (2004). The asymptotic consistency of the FRB method in multivariate regression can be shown by straightforward adaptation of the proofs of Theorem 1 and 3.

The principal purpose of the bootstrap in multivariate regression would be to obtain standard errors or confidence intervals for the individual coefficients in  $\hat{\mathcal{B}}_n$ . Analogously to the result stated in Theorem 2, Van Aelst and Willems (2005) show that the quantiles of the bootstrap distribution of  $\hat{\mathcal{B}}_n^{R*}$  have a high breakdown point, contrary to those of the classical bootstrap.

Also in Van Aelst and Willems (2005), extensive simulation results are given on the performance of the FRB confidence intervals. The latter were found to have a coverage close to the nominal value and a reasonable average length, for various data configurations. The FRB outperformed intervals based on the asymptotic variance, especially when bad leverage points were present.

Another robust approach for multivariate regression is based on robust estimates of the location and scatter of the joint distribution of the  $\mathbf{x}$ - and  $\mathbf{y}$ -variables (Rousseeuw *et al.* 2004, amongst others). If we use multivariate location S- or MM-estimates for this purpose, then the fast and robust bootstrap can readily be applied for inference in this context as well.

### 3.3 Discriminant analysis

Linear and quadratic discriminant rules are widely used for classification of multivariate data in two or more populations. For simplicity, we will restrict the discussion to the situation with two populations,  $\pi_1$  and  $\pi_2$ , having respective means  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ , and respective covariance matrices  $\Sigma_1$  and  $\Sigma_2$ . Furthermore, we will assume equal prior probabilities. The quadratic Bayes rule then classifies an observation  $\mathbf{x}$  into population  $\pi_1$  if  $d_1^Q(\mathbf{x}) > d_2^Q(\mathbf{x})$ , where

$$d_j^Q(\mathbf{x}) = -\frac{1}{2}\log|\Sigma_j| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^t \Sigma_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j); \quad j = 1, 2, \quad (36)$$

and into population  $\pi_2$  otherwise. Assuming now that  $\Sigma_1 = \Sigma_2 = \Sigma$  yields the linear rule based on the scores  $d_j^L(\mathbf{x})$ :

$$d_j^L(\mathbf{x}) = \boldsymbol{\mu}_j^t \Sigma^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_j^t \Sigma^{-1} \boldsymbol{\mu}_j; \quad j = 1, 2. \quad (37)$$

Since the true parameters  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1$  and  $\Sigma_2$  are unknown, they need to be estimated from a so-called training sample of the form  $\mathcal{Z}_n = \{\mathbf{x}_1^1, \dots, \mathbf{x}_{n_1}^1, \mathbf{x}_1^2, \dots, \mathbf{x}_{n_2}^2\} \subset \mathbb{R}^p$ . Classical linear and quadratic discriminant analysis is based on the empirical means and covariances of  $\mathcal{Z}_n$ . Alternative and more robust

methods are naturally obtained by using more robust estimates of location and scatter instead (e.g. Randles *et al.*, 1978, Hawkins and McLachlan, 1997, Croux and Dehon, 2001 and Hubert and Van Driessen, 2004). S-estimates are once again excellent candidates for this purpose.

Simply computing the S-estimates, as defined by (29), for both groups contained in  $\mathcal{Z}_n$  yields  $\hat{\boldsymbol{\mu}}_{n_1}^1$ ,  $\hat{\boldsymbol{\mu}}_{n_2}^2$ ,  $\hat{\Sigma}_{n_1}^1$  and  $\hat{\Sigma}_{n_2}^2$ , which can be plugged into the quadratic scores (36). For robust versions of the linear scores (37), where we need an estimate for the common covariance matrix  $\Sigma$ , several options exist. One possibility consists of pooling the estimates  $\hat{\Sigma}_{n_1}^1$  and  $\hat{\Sigma}_{n_2}^2$ . Another, somewhat more efficient approach is offered by He and Fung's (2000) two-sample S-estimates:  $\hat{\boldsymbol{\mu}}_n^1$ ,  $\hat{\boldsymbol{\mu}}_n^2$  and  $\hat{\Sigma}_n$  minimize  $|C|$  subject to

$$\begin{aligned} \frac{1}{n_1} \sum_{i=1}^{n_1} \rho \left( [(\mathbf{x}_i^1 - T_1)^t C^{-1} (\mathbf{x}_i^1 - T_1)]^{\frac{1}{2}} \right) + \\ \frac{1}{n_2} \sum_{i=1}^{n_2} \rho \left( [(\mathbf{x}_i^2 - T_2)^t C^{-1} (\mathbf{x}_i^2 - T_2)]^{\frac{1}{2}} \right) = b \end{aligned} \quad (38)$$

among all  $T_1, T_2 \in \mathbb{R}^p$  and  $C \in \text{PDS}(p)$ .

The performance in terms of misclassification error rates of several such robust discriminant rules has been investigated in the references mentioned above. However, the question of how to estimate the error rate in practice has not often been addressed. Hubert and Van Driessen (2004) acknowledge that the popular cross-validation (CV) technique might be too time-consuming. Therefore, they advocate to split the available data in a training set and a validation set, even though this approach is wasteful of data and has some other shortcomings. The bootstrap has often been proposed as an alternative and possible improvement on cross-validation (e.g. Efron 1983). The classical bootstrap obviously suffers from similar computational problems as cross-validation does, but since we are using S-estimates, the FRB method can be applied instead.

The idea of bootstrap error estimation is to reconstruct the discriminant rule for every bootstrap sample, based on the recalculated parameter estimates. Each time, then, the rule is evaluated on those

observations that were not included in that particular bootstrap sample and the percentage of misclassifications is computed. The averaged percentage over all bootstrap samples, say  $\widehat{\text{err}}_{\text{boot}}$ , gives an estimate of the true error rate of the original rule. This estimate generally has an upward bias and therefore Efron (1983) proposed to average it with the downwardly biased resubstitution estimate  $\widehat{\text{err}}_{\text{resub}}$  (obtained by evaluating the rule on the sample  $\mathcal{Z}_n$  itself). Efron's proposal is well-known as the .632 estimator:

$$\widehat{\text{err}}_{.632} = .632 \widehat{\text{err}}_{\text{boot}} + .368 \widehat{\text{err}}_{\text{resub}}$$

The choice of the weight .632 is based on the fact that it approximately equals the probability for a given observation to be included in a given bootstrap sample.

For the quadratic discriminant rule, the FRB formulas from Section 3.1 can simply be applied twice to obtain the recalculations  $\hat{\boldsymbol{\mu}}_{n_1}^{1R*}$ ,  $\hat{\boldsymbol{\mu}}_{n_2}^{2R*}$ ,  $\hat{\Sigma}_{n_1}^{1R*}$  and  $\hat{\Sigma}_{n_2}^{2R*}$ . For the linear discriminant rule based on the two-sample S-estimates given by (38), we can obtain analogous formulas by noting that the estimating equations for  $\hat{\boldsymbol{\theta}}_n = ((\hat{\boldsymbol{\mu}}_n^1)^t, (\hat{\boldsymbol{\mu}}_n^2)^t, \text{vec}(\hat{\Sigma}_n)^t)^t$  are given by

$$\hat{\boldsymbol{\mu}}_n^1 = \left( \sum_{i=1}^{n_1} \frac{\rho'(d_i^{(1)})}{d_i^{(1)}} \right)^{-1} \sum_{i=1}^{n_1} \frac{\rho'(d_i^{(1)})}{d_i^{(1)}} \mathbf{x}_i^1 \quad (39)$$

$$\hat{\boldsymbol{\mu}}_n^2 = \left( \sum_{i=1}^{n_2} \frac{\rho'(d_i^{(2)})}{d_i^{(2)}} \right)^{-1} \sum_{i=1}^{n_2} \frac{\rho'(d_i^{(2)})}{d_i^{(2)}} \mathbf{x}_i^2 \quad (40)$$

$$\begin{aligned} \hat{\Sigma}_n = & \frac{1}{nb} \left( \sum_{i=1}^{n_1} p \frac{\rho'(d_i^{(1)})}{d_i^{(1)}} (\mathbf{x}_i^1 - \hat{\boldsymbol{\mu}}_n^1)(\mathbf{x}_i^1 - \hat{\boldsymbol{\mu}}_n^1)^t + \left( \sum_{i=1}^{n_1} s_i^{(1)} \right) \hat{\Sigma}_n \right. \\ & \left. + \sum_{i=1}^{n_2} p \frac{\rho'(d_i^{(2)})}{d_i^{(2)}} (\mathbf{x}_i^2 - \hat{\boldsymbol{\mu}}_n^2)(\mathbf{x}_i^2 - \hat{\boldsymbol{\mu}}_n^2)^t + \left( \sum_{i=1}^{n_2} s_i^{(2)} \right) \hat{\Sigma}_n \right) \end{aligned} \quad (41)$$

where  $d_i^{(1)} = [(\mathbf{x}_i^1 - \hat{\boldsymbol{\mu}}_n^1)^t \hat{\Sigma}_n^{-1} (\mathbf{x}_i^1 - \hat{\boldsymbol{\mu}}_n^1)]^{1/2}$ ,  $s_i^{(1)} = \rho(d_i^{(1)}) - \rho'(d_i^{(1)})d_i^{(1)}$ , and analogously for  $d_i^{(2)}$  and  $s_i^{(2)}$ .

Through (6) we then once more obtain the recalculations  $\hat{\boldsymbol{\mu}}_n^{1R*}$ ,  $\hat{\boldsymbol{\mu}}_n^{2R*}$  and  $\hat{\Sigma}_n^{R*}$ .

We performed a short simulation study to examine the accuracy of the bootstrap error rate estimates in robust discriminant analysis, focusing on the linear rule based on the two-sample S-estimates. We used the same simulation design as the one reported in He and Fung (2000). That is, we have two groups of

observations, drawn from populations  $\pi_1$  and  $\pi_2$ , in dimension  $p = 3$  and we consider 5 different settings:

	$\pi_1$	$\pi_2$
A:	$50 N(0, I)$	$50 N(1, I)$
B:	$40 N(0, I) + 10 N(5, 0.25^2 I)$	$40 N(1, I) + 10 N(-4, 0.25^2 I)$
C:	$80 N(0, I) + 20 N(5, 0.25^2 I)$	$8 N(1, I) + 2 N(-4, 0.25^2 I)$
D:	$16 N(0, I) + 4 N(0, 25^2 I)$	$16 N(1, I) + 4 N(1, 25^2 I)$
E:	$58 N(0, I) + 12 N(5, 0.25^2 I)$	$25 N(1, 4I) + 5 N(-10, 0.25^2 I)$

Here,  $m N(a, S_1) + n N(b, S_2)$  means that we draw  $m$  points from  $N((a \ a \ a)^t, S_1)$  and  $n$  points from  $N((b \ b \ b)^t, S_2)$ . Hence, setting A corresponds to uncontaminated normal data, whereas in the other settings 20% of the observations are outliers. Note that He and Fung's (2000) experiment was meant to compare the rule based on two-sample S-estimates to other discriminant rules, including the classical linear rule. Here, we are not particularly concerned with the actual error rate of the rule, but rather with the performance of the bootstrap in estimating this error rate.

We generated  $N = 1000$  samples for each setting listed above, for which we computed the robust linear discriminant rule based on the 50% breakdown S-estimates (using Tukey's biweight). We then in each sample estimated the error rate of this rule through the FRB-based .632 and the classical bootstrap .632 estimator (both with varying number of bootstrap samples:  $B = 10, 30$  and  $100$ ), as well as through  $k$ -fold CV (varying  $k$ :  $k = 5, 10$  and  $n$ ) and through resubstitution. Both bootstrap and cross-validation were performed in a stratified manner, ensuring that the ratio between observations from  $\pi_1$  and from  $\pi_2$  in the new samples is approximately similar to the ratio in the original sample  $\mathcal{Z}_n$ . Also, observations that were identified as outliers through their Mahalanobis distances based on the two-sample S-estimates, were always ignored when computing the misclassification percentages.

**Table 4** Root mean squared error for misclassification rate estimates for linear discriminant rule based on two-sample maximum breakdown S-estimates

		A	B	C	D	E	time
$\widehat{\text{error}}_{.632}$ (FRB)	$B = 10$	.045	.047	.067	.076	.094	.03
	$B = 30$	.044	.045	.065	.072	.093	.06
	$B = 100$	.044	.045	.064	.070	.093	.16
$\widehat{\text{error}}_{.632}$ (Classical)	$B = 10$	.046	.055	.065	.080	.091	1.37
	$B = 30$	.045	.051	.062	.076	.089	4.12
	$B = 100$	.045	.049	.062	.074	.088	13.72
$\widehat{\text{error}}_{\text{CV}}$	$k = 5$	.048	.053	.067	.092	.094	.64
	$k = 10$	.048	.051	.068	.086	.096	1.34
	$k = n$	.048	.048	.068	.085	.097	13.80
$\widehat{\text{error}}_{\text{resub}}$		.053	.050	.072	.093	.117	.00
“True” error rate		.204	.205	.215	.223	.290	

Table 4 lists the root mean squared error of the different estimates for the misclassification rate. The “true” error rate is obtained by evaluating the rules on a large test sample ( $n = 2000$ ) drawn from the uncontaminated parts of  $\pi_1$  and  $\pi_2$  respectively, and is also reported in Table 4. The last column contains the corresponding average computation times (in CPU seconds) for a sample from setting A. We used a standard implementation of Salibián-Barrera and Yohai’s (2006) algorithm for S-estimators, adapted for the multivariate two-sample situation. These values are mainly listed for comparison reasons.

The mean squared errors (MSEs) seem to indicate that both the FRB and the classical bootstrap estimates on average have a slight advantage over CV. The two bootstrap methods take turns in achieving



the lowest MSE values for the five situations, with the differences being relatively small. The FRB estimates of course have the advantage of being much faster to compute. It should be noted that the difference in computation time will become more of practical significance when the sample sizes become larger than in this experiment.

Concerning CV, the possibly very time-consuming  $n$ -fold CV does not always yield better results than the faster 5-fold CV. Note, however, that even 5-fold CV is much slower than the FRB and that small values of  $k$  generally carry a higher risk of numerical instability due to high percentages of outliers, similarly to the classical bootstrap situation. In this simulation study, though, this type of numerical instability did not occur in CV nor in the classical bootstrap: the 20% outliers in the original sample did not constitute a real challenge to the 50% breakdown point of the S-estimator.

Table 4 further shows that increasing the number of bootstrap samples  $B$  generally improves the accuracy of the error rate estimates, as could be expected. Given the small computational cost of the FRB method, it seems reasonable to choose a relatively high number, say  $B = 100$  at least. Note, finally, that the .632 estimates for every setting outperform the resubstitution estimate.

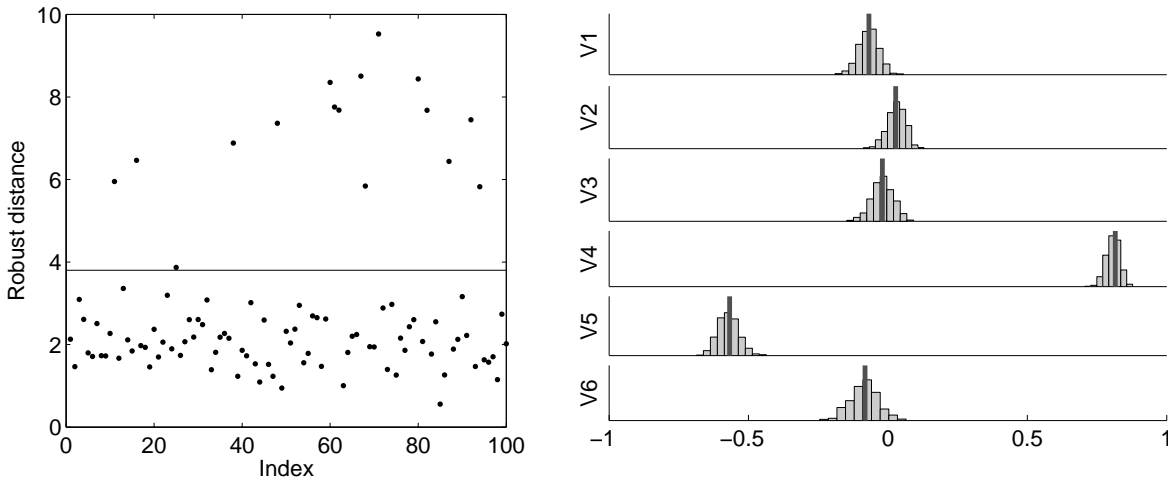
## 4 Examples

In this section we give two real data examples to illustrate the use and the benefits of the FRB method in the context of PCA and discriminant analysis, respectively. For examples in the context of linear regression, we refer to Salibian-Barrera (2000), Salibian-Barrera and Zamar (2002) and Van Aelst and Willems (2005).

#### 4.1 PCA : Bank notes data

For an illustration of the FRB method for robust PCA, we consider the measurements on  $n = 100$  forged old Swiss 1000 franc bills, which are part of the ‘Swiss bank notes data’ from Flury and Riedwyl (1988). The data consist of  $p = 6$  variables corresponding to length, height and other distance measurements on the bills, and was also analysed in Salibián-Barrera *et al.* (2006). We applied PCA based on the 50% breakdown S-estimator. The distances  $d_i$  based on the robust location and covariance estimates are shown in Figure 1 (left). It can be seen that the S-estimator detected a group of 15 outlying observations (with large robust distances).

The S-estimates for the eigenvalues of the shape matrix are given by  $\hat{\lambda} = [1.38, 0.24, 0.14, 0.07, 0.06, 0.03]$ . The coefficients or *weights* of the first PC are  $\hat{v}_1 = [-0.068 \ 0.028 \ -0.020 \ 0.816 \ -0.568 \ -0.082]$ . Hence, the first principal component can be interpreted as the difference between the 4th and the 5th



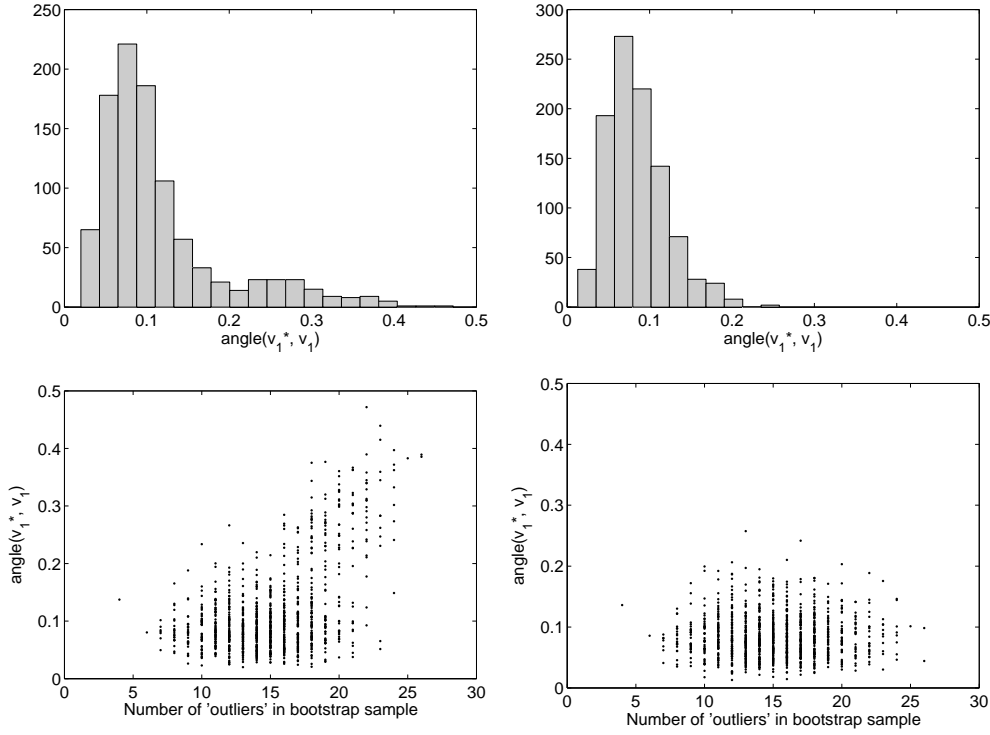
**Fig. 1** Swiss bank notes data; Left: robust distances based on S; Right: weights in the first PC, with FRB histograms

variable. These correspond to the distance from the inner frame on the bill to respectively the lower border and the upper border.

The right panel of Figure 1 plots the weights in  $\hat{v}_1$  along with bootstrap histograms obtained by performing the FRB with  $B = 1000$ . This is another application of bootstrap for PCA, not mentioned in the previous section. Some care is needed here, since the coefficients of the principal components are not uniquely defined. In order to obtain meaningful inference results, we imposed that, for each  $\hat{v}_j; j = 1, \dots, 6$ , the coefficient with the largest absolute value should be positive in every bootstrap recalculation, as well as in the S-estimate  $\hat{v}_j$  itself. The bootstrap result in Figure 1 indicates that the coefficients of  $\hat{v}_1$ , i.e. the weights associated with the first PC, are quite stable. Note that the bootstrap can also be used to construct confidence intervals for the weights and to determine which original components contribute significantly to a PC.

As an alternative way of assessing the stability of the first PC, we look at the bootstrap distribution of the angles  $\arccos(|\hat{v}_1^* \hat{v}_1|)$ . The upper panels in Figure 2 compare the result from the classical bootstrap (left) with that from the FRB (right).

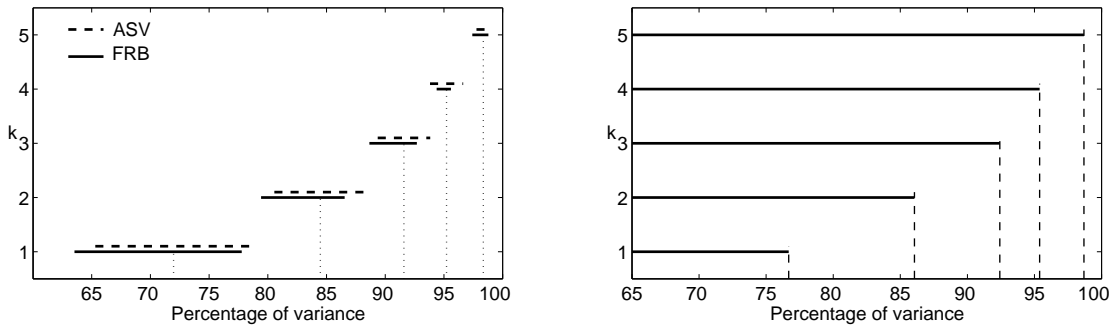
While the fast and robust bootstrap recalculations yield angles no larger than about 0.2, the classical bootstrap suggests a somewhat higher variability. This is in fact an illustration of the instability of the classical bootstrap. Indeed, it turns out that the bootstrap samples corresponding to the larger angles all contained more than 15 replications from the original group of 15 outliers. This is shown in the lower panels in Figure 2, where we plotted for each bootstrap sample the angle between  $\hat{v}_1$  and  $\hat{v}_1^*$  versus the total number of replications of the 15 outliers in that bootstrap sample. The left panel again corresponds to the classical bootstrap, the right to the fast and robust bootstrap. Clearly, in the classical procedure higher proportions of outlying observations give rise to a larger variability concerning the eigenvector



**Fig. 2** Swiss bank notes data: Upper: bootstrap histograms for the angles between  $\hat{v}_1^*$  and  $\hat{v}_1$ ; Lower: angles between  $\hat{v}_1^*$  and  $\hat{v}_1$  versus number of observations from the outlier group present in the bootstrap sample; Left: classical bootstrap; Right: FRB

angles. On the other hand, in the FRB procedure the group of 15 outliers is severely downweighted, due to the large robust distances, and hence has very little influence on the recalculated eigenvectors.

We now turn to the question of how many components should be retained. The S-estimates yield  $\hat{p}_1 = 72.0\%$ ,  $\hat{p}_2 = 84.5\%$ ,  $\hat{p}_3 = 91.6\%$ ,  $\hat{p}_4 = 95.2\%$  and  $\hat{p}_5 = 98.4\%$ . We used the FRB to construct 95% confidence intervals for these percentages. The left panel of Figure 3 shows the bootstrap intervals (BCa method; see Davison and Hinkley, 1997, page 202) and as a comparison also gives the intervals based on an empirical version of the asymptotic variance. The classical bootstrap intervals, which are not shown on the plot, were very similar to the FRB intervals. The dotted vertical lines indicate the point



**Fig. 3** Swiss bank notes data. Left: 95% confidence intervals for the percentage of variance explained by the first  $k$  principal components; FRB (solid) compared to intervals based on asymptotic normality (dashed). Right: 95% one-sided confidence intervals (FRB) for the percentage of variance explained by the first  $k$  principal components

estimates. The difference between both methods here is rather small, although the intervals based on asymptotic normality are slightly more optimistic. Recall that the simulation study in Salibián-Barrera *et al.* (2006) demonstrated that bootstrap intervals usually have better coverage. The intervals in general are somewhat long, although there is no overlap present. Concerning the number of PCs to retain, suppose we adopt the criterion of choosing the smallest  $k$  for which the percentage  $p_k$  exceeds some cutoff value. Then, if the cutoff is set at 70%, we might decide to play it safe and take two components into account, instead of just the first. Indeed, the point estimate  $\hat{p}_1$  equals 72.0%, but the lower limit of the confidence interval is as low as 63.5%. If the cutoff is set at 80%, the choice for  $k = 2$  would be obvious since both the point estimate and almost the whole interval exceeds 80%. In case of a 90% cutoff, the estimate  $\hat{p}_3$  leads to the choice of  $k = 3$  but the corresponding interval again suggests to be careful and to choose  $k = 4$  instead.

Finally, note that we could apply the duality between hypothesis tests and confidence intervals to test null hypotheses of the form  $H_0 : p_k \geq \pi$ . Here,  $\pi$  would then equal e.g. 80% or 90%. We can decide to accept  $H_0$  at the level  $\alpha$  whenever  $\pi$  is contained in a  $(1 - \alpha)100\%$  one-sided confidence interval for

$p_k$ . As an example we plotted in the right panel of Figure 3 the 95% one-sided intervals for  $p_k$  based on the fast bootstrap, where we again used the BCa method. On this plot we can immediately see which hypotheses would be accepted on the 5% significance level. E.g.  $H_0 : p_k \geq 80\%$  is accepted for  $k \geq 2$ , while  $H_0 : p_k \geq 90\%$  is accepted for  $k \geq 3$ .

#### 4.2 Discriminant analysis: Hemophilia data

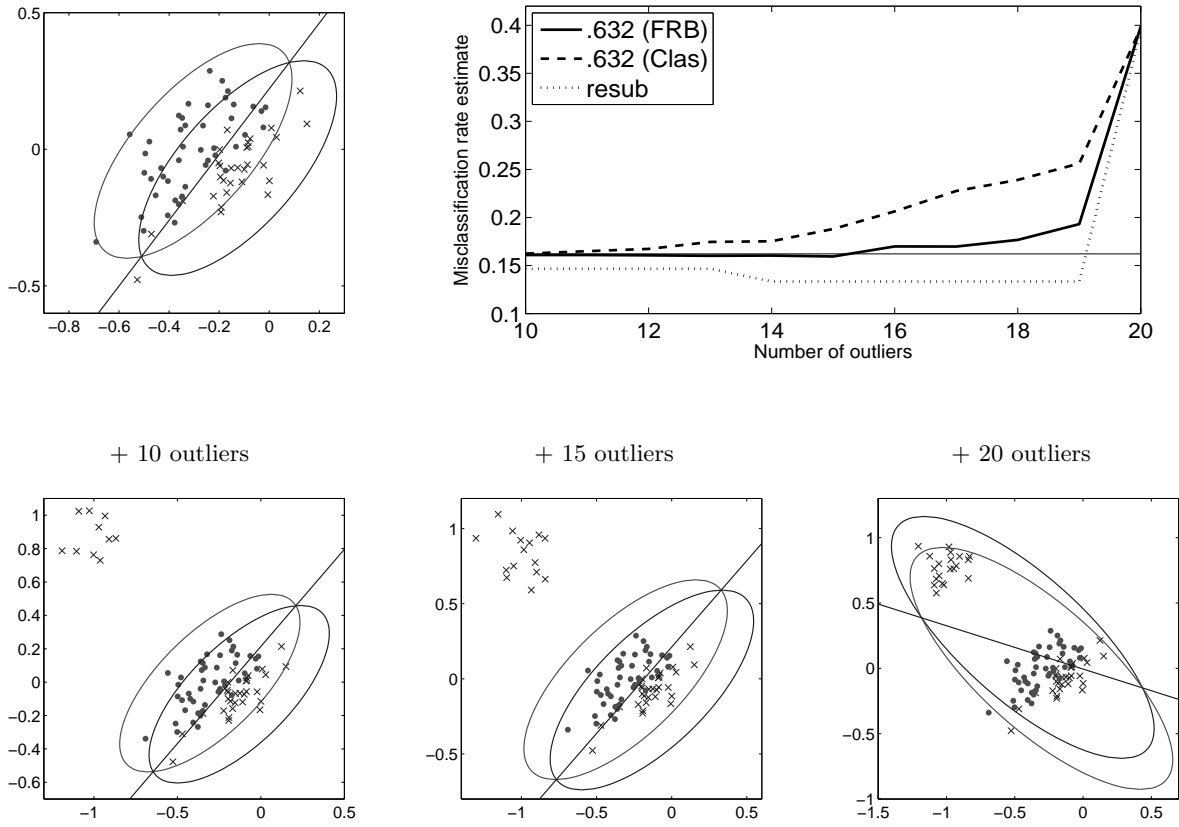
For an example in the context of robust discriminant analysis we turn to the Hemophilia data (Habbema *et al.* 1974), which consists of  $n_1 = 30$  observations on normal women and  $n_2 = 45$  on hemophilia A carriers, in  $p = 2$  variables. The data are depicted in the upper left panel of Figure 4. Group 1 observations are plotted by crosses (x), group 2 observations by dots (•). Robust discriminant analysis methods were applied to these data by Hawkins and McLachlan (1997) and by Hubert and Van Driessen (2004), both of which concluded that their robust method gave similar results as the classical discriminant method. Such a conclusion can be considered desirable for a robust method in this case, since it is clear that there are no real outliers in the data and hence the classical method's performance would be close to optimal.

Here we applied the robust linear discriminant method based on S-estimates, where we chose to tune the biweight function such that we have a 25% breakdown value (and hence a better efficiency than the 50% breakdown version). The resulting discriminant line is again very similar to the classical line. It is superimposed on the data in the upper left panel of Figure 4, which also shows the tolerance ellipses of the two groups. The location and shape of the tolerance ellipses are based on the two-sample S-estimates  $(\hat{\mu}_n^1, \hat{\Sigma}_n)$  and  $(\hat{\mu}_n^2, \hat{\Sigma}_n)$  respectively. The volumes are such that the ellipses include those observations from the group with distances  $(d_i^{(1)}$  resp.  $d_i^{(2)})$  not larger than the .975  $\chi_p^2$  quantile. It can be seen that only a few very small outliers (i.e. observations outside their respective tolerance ellipse) are found by the S-estimates.

To estimate the misclassification rate of the discriminant rule based on the S-estimates, we applied both the FRB and the classical bootstrap, with  $B = 100$ . Averaged over 10 runs (corresponding to different bootstrap samples), the .632 rules of the two bootstrap procedures respectively estimated .1621 and .1625. Hence, the FRB and the classical bootstrap did not yield appreciably different results here, except that the FRB was much faster. This indicates that the FRB is not only fast and robust but also accurately mimics the classical bootstrap whenever the latter is supposed to work well, i.e. in outlier-free data.

Now to get an appreciation of the robustness of the FRB method, we did a small experiment by drawing observations from  $N_2((-1, .8)^t, .1\mathbf{I})$ , sequentially adding them to the data and labeling them as belonging to group 1. This results in the contaminated data depicted in the lower panels of Figure 4, for 10, 15 and 20 outliers respectively. These plots also show the results of the linear discriminant rule based on the S-estimates. As can be seen, the S-estimates are not influenced much by the outliers until the number of outliers reaches 20. At this point the S-estimator no longer detects and downweights the outliers, and the method breaks down.

While the S-estimates may not be bothered by less than 20 such outliers, it is interesting to ask whether the corresponding bootstrap error rate estimates remain unaffected as well. The upper right panel of Figure 4 gives the answer, by plotting the error rate estimates as a function of the number of outliers, starting at 10. The solid line corresponds to the FRB, the dashed line to the classical bootstrap, and the dotted line are the resubstitution estimates. The bootstrap estimates are again averaged over 10 runs. The horizontal line indicates the error rate as estimated on the original data (.162). The FRB estimates deviate very little from the original estimate, until the S-estimates themselves finally break down at 20 outliers. On the other hand, the classical bootstrap estimates are clearly affected by the outliers, starting from 15 outliers say, long before the discriminant rule itself is affected. This can be



**Fig. 4** Upper left: Hemophilia data with S-based linear discriminant result; Upper right: misclassification rate estimates as a function of the number of added outliers; Lower: Hemophilia data contaminated with 10, 15, and 20 outliers respectively, with S-based linear discriminant result

attributed to the fact that some bootstrap samples will contain enough outliers to break down the S-estimates, resulting in many misclassifications corresponding to those bootstrap samples and thus an overestimation of the error rate.



## 5 Conclusion

The fast and robust bootstrap method reviewed in this paper has been proven to be a practical and reliable tool for inference based on robust estimators in various settings. The method is especially valuable in linear regression, for which robust inference in terms of confidence intervals and hypothesis tests is of crucial importance. Its application, however, is not limited to robust regression. It has been demonstrated that the FRB performs well as an inference technique in multivariate models, particularly regarding robust principal components analysis. Also, in this paper we have given an example of the method's usefulness in robust discriminant analysis.

The FRB procedure was specifically developed for S- and MM-estimators, which are among the most prominent robust estimators in the literature as well as in practice. Other estimators to which the method can be applied include GM-estimators and  $\tau$ -estimators.

## References

1. Adrover, J., Salibian-Barrera, M. and Zamar, R. (2004). Globally robust inference for the location and simple linear regression models. *Journal of Statistical Planning and Inference*, **119**, 353–375.
2. Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147–185.
3. Beran, R. and Srivastava, M.S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, **13**, 95–115.
4. Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.
5. Campbell, N.A. (1980). Robust procedures in multivariate analysis: robust covariance estimation. *Applied Statistics*, **29**, 231–237.
6. Carroll, R.J. (1978). On almost sure expansions for M-estimates. *The Annals of Statistics*, **6**, 314–318.

7. Carroll, R.J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *Journal of the American Statistical Association*, **74**, 674–679.
8. Carroll, R. J. and Welsh, A. H. (1988). A Note on Asymmetry and Robustness in Linear Regression. *The American Statistician*, **42**, 285–287.
9. Croux, C. and Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *Canadian Journal of Statistics*, **29**, 473–492.
10. Croux, C., Dhaene, G. and Hoorelbeke, D. (2003). Robust standard errors for robust estimators. Research report, Dept. of Applied Economics, K.U. Leuven.
11. Croux, C. and Haesbroeck, G. (2000). Principal components analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, **87**, 603–618.
12. Davies, P.L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, **15**, 1269–1292.
13. Davies, P.L. (1990). The asymptotics of S-estimators in the linear regression model. *The Annals of Statistics*, **18**, 1651–1675.
14. Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
15. Daudin, J.J., Duby, C. and Trecourt, P. (1988). Stability of principal component analysis studied by the bootstrap method. *Statistics*, **19**, 241–258.
16. Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, **76**, 354–362.
17. Diaconis, P. and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, **248**, 96–108.
18. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**, 1–26.
19. Efron, B. (1983). Estimating the error rate of a prediction rule: some improvements on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.

20. Fisher, N.I. and Hall, P. (1990). On bootstrap hypothesis testing. *The Australian Journal of Statistics*, **32**, 177–190.
21. Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach*, Cambridge University Press.
22. Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, **9**, 1218–1228.
23. Habbema, J.D.F., Hermans, J. and Van den Broeck, K. (1974). A stepwise discriminant analysis program using density estimation. In: Bruckmann, G., Ferschl, F. and Schmetterer, L. (Eds.) *Proceedings in Computational Statistics*, Physica-Verlag, Vienna, pp. 101–110.
24. Hall, P. and Wilson, S.R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, **47**, 757–762.
25. Hawkins, D.M. and McLachlan, G.J. (1997). High breakdown linear discriminant analysis. *Journal of the American Statistical Association*, **92**, 136–143.
26. He, X. and Fung, W.K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis*, **72**, 151–162.
27. Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.
28. Hubert, M. and Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, **45**, 301–320.
29. Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded influence regression estimation. *Journal of the American Statistical Association*, **77**, 595–604.
30. Lopuhaä, H. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, **17**, 1662–1683.
31. Lopuhaä, H. (1992). Multivariate tau-estimators. *Canadian Journal of Statistics*, **19**, 307–321.
32. Markatou, M., Stahel, W. and Ronchetti, E. (1991). Robust M-type testing procedures for linear models. In *Directions in robust statistics and diagnostics, Part I*, pp. 201–220, IMA Vol. Math. Appl., 33, Springer, New York.
33. Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, **4**, 51–67.

34. Omelka, M. and Salibian-Barrera, M. (2006). Uniform asymptotics for S- and MM-regression estimators. Unpublished manuscript. Available on-line at <http://hajek.stat.ubc.ca/~matias/pubs.html>.
35. Randles, R.H., Broffitt, J.D., Ramsberg, J.S. and Hogg, R.V. (1978). Generalized linear and quadratic discriminant functions using robust estimators. *Journal of the American Statistical Association*, **73**, 564–568.
36. Rocke, D.M. and Downs, G.W. (1981). Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap. *Communications in Statistics, Part B – Simulation and Computation*, **10**, 221–248.
37. Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
38. Rousseeuw, P.J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics, Part A - Theory and Methods*, **22**, 965–984.
39. Rousseeuw, P.J., Van Aelst, S., Van Driessen, K. and Agullo, J. (2004). Robust multivariate regression. *Technometrics*, **46**, 293–305.
40. Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series*. (J. Franke, W. Hardle and D. Martin, eds.). *Lecture Notes in Statist.*, **26** 256-272. Berlin: Springer-Verlag.
41. Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. Unpublished Ph.D. Thesis. University of British Columbia, Department of Statistics, Vancouver, BC. Available on-line at <ftp://ftp.stat.ubc.ca/pub/matias/Thesis>.
42. Salibian-Barrera, M. (2005). Estimating the p-values of robust tests for the linear model. *Journal of Statistical Planning and Inference*, **128**, 241–257.
43. Salibian-Barrera, M. (2006a). The asymptotics of MM-estimators for linear regression with fixed designs. *Metrika*, **63**, 283-294.
44. Salibian-Barrera, M. (2006b). Bootstrapping MM-estimators for linear regression with fixed designs. *Statistics and Probability Letters*, **76**, 1287-1297.
45. Salibian-Barrera, M., Van Aelst, S. and Willems, G. (2006). PCA based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, **101**, 1198–1211.

46. Salibian-Barrera, M. and Yohai, V. (2006). A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics*, **15**, 414–427.
47. Salibian-Barrera, M. and Zamar, R. H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, **30**, 556–582.
48. Singh, K. (1998). Breakdown theory for bootstrap quantiles. *The Annals of Statistics*, **26**, 1719–1732.
49. Stromberg, A.J. (1997). Robust covariance estimates based on resampling. *Journal of Statistical Planning and Inference*, **57**, 321–334.
50. Tatsuoka, K.S. and Tyler, D.E. (2000). The uniqueness of S and M-functionals under non-elliptical distributions. *The Annals of Statistics*, **28**, 1219–1243.
51. Van Aelst, S. and Willems, G. (2004). Multivariate regression S-estimators for robust estimation and inference. Technical report, available at <http://users.ugent.be/~svaelst/publications.html>.
52. Van Aelst, S. and Willems, G. (2005). Multivariate regression S-estimators for robust estimation and inference. *Statistica Sinica*, **15**, 981–1001.
53. Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642–656.
54. Yohai, V.J. and Zamar, R.H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, **83**, 406–413.
55. Yohai, V.J. and Zamar, R.H. (1998). Optimal locally robust M-estimates of regression. *Journal of Statistical Planning and Inference*, **64**, 309–323.